

Abstract

Baseball Player Analysis-The Statcast revolution

Statcast is a state-of-the-art tracking system that uses high-resolution cameras and radar equipment to measure the precise location and movement of baseballs and baseball players. Introduced in 2015 to all 30 major league ballparks, Statcast data is revolutionizing the game. In this project , we're going to wrangle, analyse, and visualize Statcast data to compare Mr. Judge and another (extremely large) teammate of his- Giancarlo Stanton. There are two datasets given, which contain Statcast data of two players for 2015-2017. We'll use pandas Data Frames to store this data and by using libraries, matplotlib and seaborn we will visualize the data.

Tools / Skills Used

1. Python Programming
2. Jupyter Notebook
3. Pandas
4. NumPy
5. Matplotlib
6. Seaborn
7. Exploratory Data Analysis
8. Data Visualization

Problem Statement - Baseball Player Analysis-The Statcast revolution

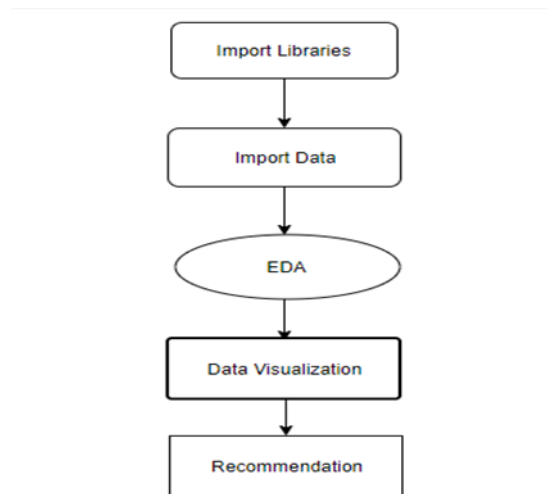
Analyse Player's Statcast Data:

Statcast is a state-of-the-art tracking system that uses high-resolution cameras and radar equipment to measure the precise location and movement of baseballs and baseball players. Introduced in 2015 to all 30 major league ballparks, Statcast data is revolutionizing the game. Teams are engaging in an "arms race" of data analysis, hiring analysts left and right in an attempt to gain an edge over their competition.

In this project, we're going to wrangle, analyse, and visualize Statcast data to compare Mr. Judge and another (extremely large) teammate of his- Giancarlo Stanton. There are two datasets given, which contain Statcast data of two players for 2015-2017.

Implementation

Workflow:



Key points:

- 1. Statcast Measure:** Statcast is capable of measuring the exit velocity, launch angle and vector of the ball as it comes off the bat. From there, Statcast can also track the hang time and projected distance that a ball travels.
- 2. Player Profile:** Giancarlo Stanton is also a very large human being, standing 6 feet 6 inches tall and weighing 245 pounds. Judge and Stanton will be teammates on the New York Yankees. They are similar in a lot of ways, one being that they hit a lot of home runs. Stanton and Judge led baseball in home runs in 2017, with 59 and 52 , respectively. These are exceptional totals - the player in third "only" had 45 home runs. Stanton and Judge are also different in many ways. One is batted ball events, which is any batted ball that produces a result. This includes outs, hits, and errors.

Code Snippets:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

# Load Aaron Judge's Statcast data
judge= pd.read_csv('C:/Users/Lenovo/Desktop/Capstone project/A-New-Era-for-Data-Analysis-in-Baseball-MLB-Statcast--master/A New Era')

# Load Giancarlo Stanton's Statcast data
stanton=pd.read_csv('C:/Users/Lenovo/Desktop/Capstone project/A-New-Era-for-Data-Analysis-in-Baseball-MLB-Statcast--master/A New Era')
```

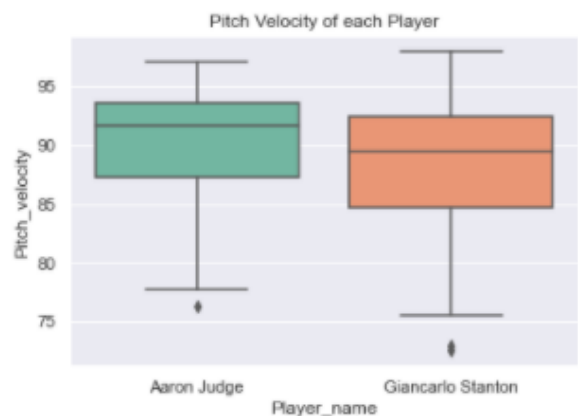
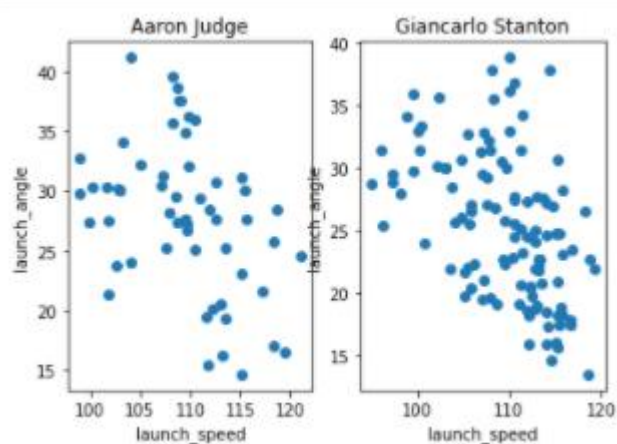
7. Home runs by pitch location (II)

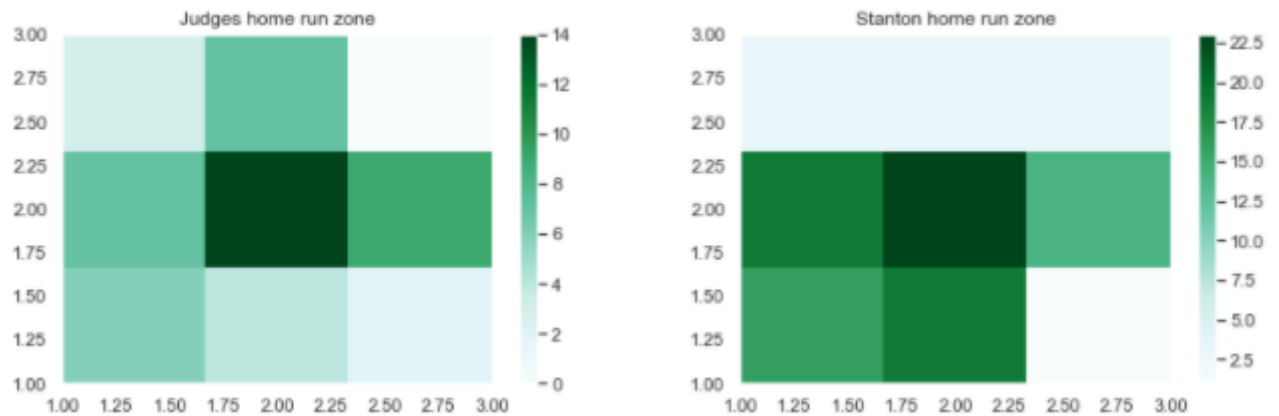
And let's do the same but for y-coordinates.

```
def assign_x_coord(loc):
    """
    Assigns a y-coordinate to Statcast's strike zone numbers. Zones 11, 12, 13,
    and 14 are ignored for plotting simplicity.
    """
    # Left third of strike zone-1
    if loc.zone in [1, 4, 7]:
        return 1
    # Middle third of strike zone-2
    if loc.zone in [2, 5, 8]:
        return 2
    # Right third of strike zone-3
    if loc.zone in [3, 6, 9]:
        return 3

def assign_y_coord(loc):
    """
    Assigns a y-coordinate to Statcast's strike zone numbers. Zones 11, 12, 13,
    and 14 are ignored for plotting simplicity.
    """
    # Upper third of strike zone
    if loc.zone in [1, 2, 3]:
        return 3
    # Middle third of strike zone
    if loc.zone in [4, 5, 6]:
        return 2
    # Lower third of strike zone
    if loc.zone in [7, 8, 9]:
        return 1
```

Visualization Snippets:





Conclusion/ Results

- Stanton does not hit many home runs on pitches in the upper third of the strike zone.
- Like pretty much every hitter ever, both players love pitches in the horizontal and vertical middle of the plate.
- Judge's least favourite home run pitch appears to be high-away while Stanton's appears to be below-away.
- If we were to describe Stanton's home run zone, it'd be middle-inside. Judge's home run zone is much more spread out.

The grand takeaway from this whole exercise: Aaron Judge and Giancarlo Stanton are not identical despite their superficial similarities. In terms of home runs, their launch profiles, as well as their pitch speed and location preferences, are different. file. Hence, **opposing pitchers should be still scared.**

Future Scope

In future, Statcast Data can be used to compare player from one team to another to get their weakness and to analyse more accurate results.